BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 7: e113053
doi: 10.3897/biss.7.113053

OPEN ACCESS

Conference Abstract

No Pain No Gain: Standards mapping in Latimer Core development

Matt Woodburn[‡], Jutta Buschbom[§], Sharon Grant[|], Janeen Jones[|], Ben Norton[¶], Maarten Trekels[#], Sarah Vincent[‡], Kate Webbink[¤]

‡ Natural History Museum, London, United Kingdom
§ Statistical Genetics, Ahrensburg, Germany
| Field Museum of Natural History, Chicago, Illinois, United States of America
¶ Ecosystem Planning and Restoration, Raleigh, North Carolina, United States of America
# Meise Botanic Garden, Meise, Belgium
¤ Field Museum of Natural History, Chicago, United States of America

Corresponding author: Matt Woodburn (m.woodburn@nhm.ac.uk)

Received: 20 Sep 2023 | Published: 21 Sep 2023

Citation: Woodburn M, Buschbom J, Grant S, Jones J, Norton B, Trekels M, Vincent S, Webbink K (2023) No Pain No Gain: Standards mapping in Latimer Core development. Biodiversity Information Science and Standards 7: e113053. https://doi.org/10.3897/biss.7.113053

Abstract

Latimer Core (LtC) is a new proposed Biodiversity Information Standards (TDWG) data standard that supports the representation and discovery of natural science collections by structuring data about the groups of objects that those collections and their subcomponents encompass (Woodburn et al. 2022). It is designed to be applicable to a range of use cases that include high level collection registries, rich textual narratives and semantic networks of collections, as well as more granular, quantitative breakdowns of collections to aid collection discovery and digitisation planning.

As a standard that is (in this first version) focused on natural science collections, LtC has significant intersections with existing data standards and models (Fig. 1) that represent individual natural science objects and occurrences and their associated data (e.g., Darwin Core (DwC), Access to Biological Collection Data (ABCD), Conceptual Reference Model of the International Committee on Documentation (CIDOC-CRM)). LtC's scope also overlaps with standards for more generic concepts like metadata, organisations, people and activities (i.e., Dublin Core, World Wide Web Consortium (W3C) ORG Ontology and PROV Ontology, Schema.org). LtC represents just an element of this extended network of data

This is an open access article distributed under the terms of the CC0 Public Domain Dedication.

standards for the natural sciences and related concepts. Mapping between LtC and intersecting standards is therefore crucial for avoiding duplication of effort in the standard development process, and ensuring that data stored using the different standards are as interoperable as possible in alignment with FAIR (Findable, Accessible, Interoperable, Reusable) principles. In particular, it is vital to make robust associations between records representing groups of objects in LtC and records (where available) that represent the objects within those groups.

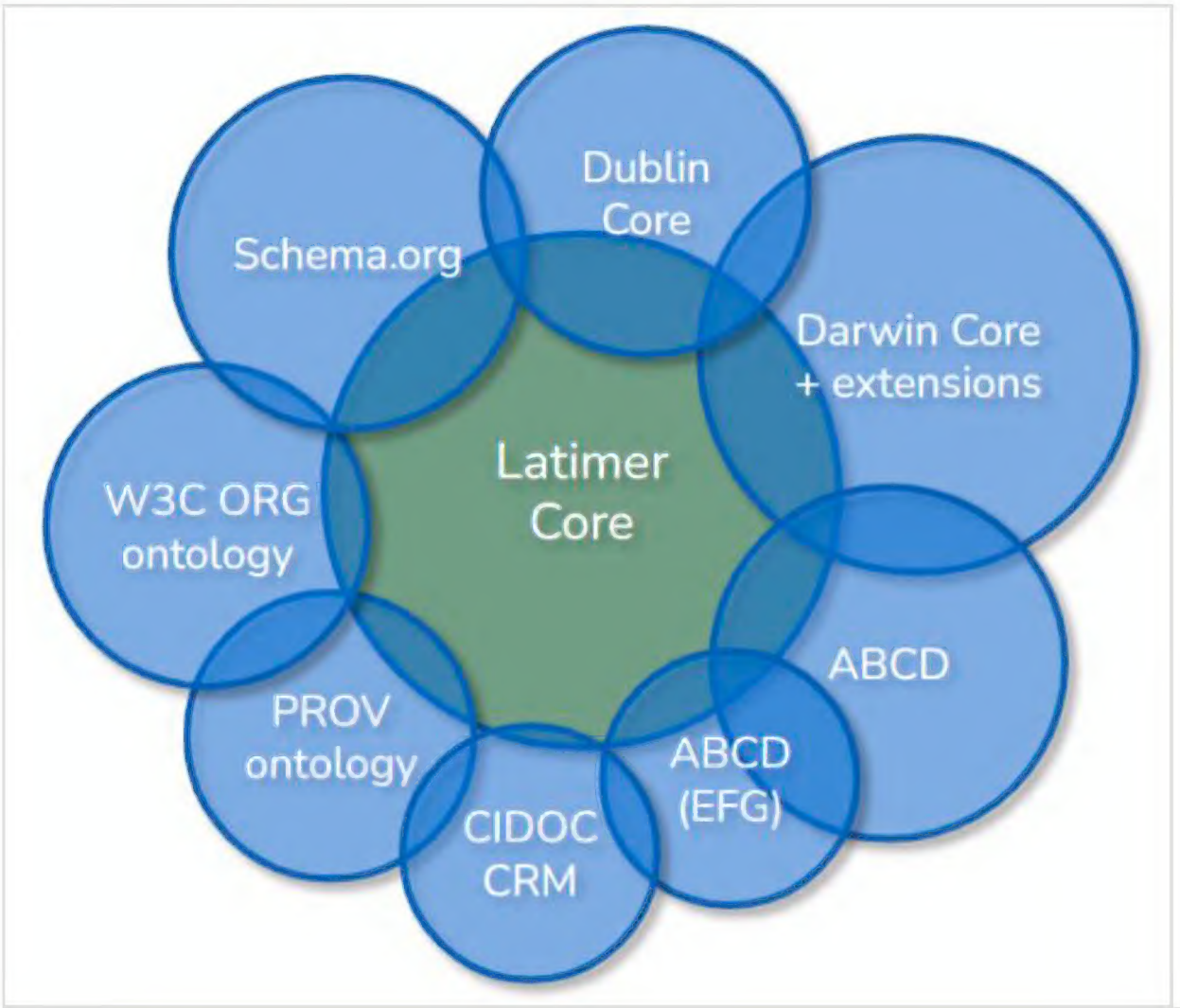

Figure 1.

A visual representation of some of the standards with which LtC intersects.

During LtC development, efforts were made to identify and align with relevant standards and vocabularies, and adopt existing terms from them where possible. During expert review, a more structured approach was proposed and implemented using the Simple Knowledge Organization System (SKOS) mappingRelation vocabulary. This exercise helped to better describe the nature of the mappings between new LtC terms and related terms in other standards, and to validate decisions around the borrowing of existing terms for LtC. A further exercise also used elements of the Simple Standard for Sharing Ontological Mappings (SSSOM) to start to develop a more comprehensive set of metadata around these mappings. At present, these mappings (Suppl. material 1 and Suppl. material 2) are provisional and not considered to be comprehensive, but should be further refined and expanded over time.

Even with the support provided by the SKOS and SSSOM standards, the LtC experience has proven the mapping process to be far from straightforward. Different standards vary in how they are structured, for example, DwC is a ‘bag of terms’, with informal classes and no structural constraints, while more structured standards and ontologies like ABCD and PROV employ different approaches to how structure is defined and documented. The various standards use different metadata schemas and serialisations (e.g., Resource Description Framework (RDF), XML) for their documentation, and different approaches to providing persistent, resolvable identifiers for their terms. There are also many subtle nuances involved in assessing the alignment between the concepts that the source and target terms represent, particularly when assessing whether a match is exact enough to allow the existing term to be adopted. These factors make the mapping process quite manual and labour-intensive. Approaches and tools, such as developing decision trees (Fig. 2) to represent the logic involved and further exploration of the SSSOM standard, could help to streamline this process.

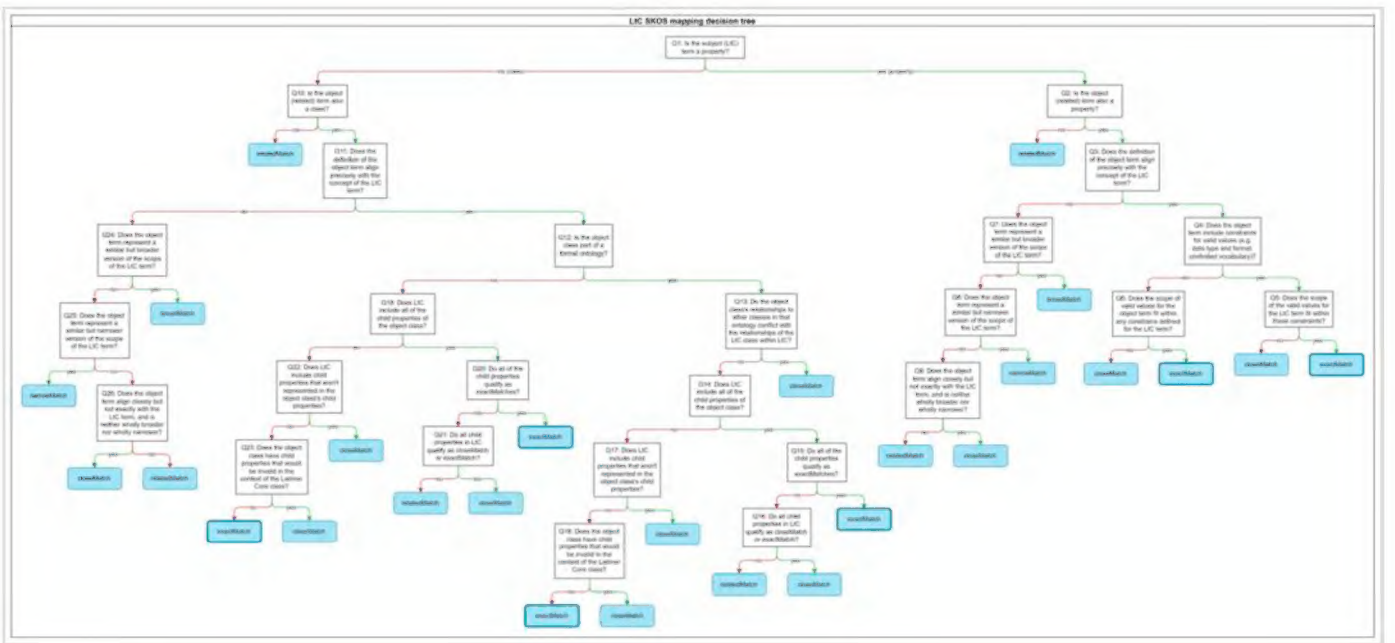

Figure 2.

A provisional decision tree created to structure and streamline the process of defining mappings between LtC and other standards.

In this presentation, we will discuss the LtC experience of the standard mapping process, the challenges faced and methods used, and the potential to contribute this experience to a collaborative standards mapping within the anticipated TDWG Standards Mapping Interest Group.

Keywords

collection descriptions, SKOS, SSSOM

Presenting author

Matt Woodburn

Presented at

TDWG 2023

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Woodburn M, Buschbom J, Droege G, Grant S, Groom Q, Jones J, Trekels M, Vincent S, Webbink K (2022) Latimer Core: A new data standard for collection descriptions. Biodiversity Information Science and Standards 6 https://doi.org/10.3897/biss.6.91159

Supplementary materials

Suppl. material 1: SKOS mappings doi

Authors: Matt Woodburn
Data type: SKOS term mappings
Brief description: Provisional SKOS mappings between LtC terms and related terms in other standards.
Download file (14.11 kb)

Suppl. material 2: Extended SSSOM mappings doi

Authors: Matt Woodburn, Jutta Buschbom
Data type: SSSOM term mappings
Brief description: Provisional extended mappings using elements of the SSSOM schema.
Download file (96.99 kb)